JAMA Psychiatry | Original Investigation

Development and Validation of a Model to Predict Posttraumatic Stress Disorder and Major Depression After a Motor Vehicle Collision

Hannah N. Ziobrowski, PhD, MPH; Chris J. Kennedy, PhD; Berk Ustun, PhD; Stacey L. House, MD, PhD; Francesca L. Beaudoin, MD, PhD; Xinming An, PhD; Donglin Zeng, PhD; Kenneth A. Bollen, PhD; Maria Petukhova, PhD; Nancy A. Sampson, BA; Victor Puac-Polanco, MD, DrPH; Sue Lee, MA; Karestan C. Koenen, PhD; Kerry J. Ressler, MD, PhD; Samuel A. McLean, MD, MPH; Ronald C. Kessler, PhD; and the AURORA Consortium

IMPORTANCE A substantial proportion of the 40 million people in the US who present to emergency departments (EDs) each year after traumatic events develop posttraumatic stress disorder (PTSD) or major depressive episode (MDE). Accurately identifying patients at high risk in the ED would facilitate the targeting of preventive interventions.

OBJECTIVES To develop and validate a prediction tool based on ED reports after a motor vehicle collision to predict PTSD or MDE 3 months later.

DESIGN, SETTING, AND PARTICIPANTS The Advancing Understanding of Recovery After Trauma (AURORA) study is a longitudinal study that examined adverse posttraumatic neuropsychiatric sequalae among patients who presented to 28 US urban EDs in the immediate aftermath of a traumatic experience. Enrollment began on September 25, 2017. The 1003 patients considered in this diagnostic/prognostic report completed 3-month assessments by January 31, 2020. Each patient received a baseline ED assessment along with follow-up self-report surveys 2 weeks, 8 weeks, and 3 months later. An ensemble machine learning method was used to predict 3-month PTSD or MDE from baseline information. Data analysis was performed from November 1, 2020, to May 31, 2021.

MAIN OUTCOMES AND MEASURES The PTSD Checklist for *DSM-5* was used to assess PTSD and the Patient Reported Outcomes Measurement Information System Depression Short-Form 8b to assess MDE.

RESULTS A total of 1003 patients (median [interquartile range] age, 34.5 [24-43] years; 715 [weighted 67.9%] female; 100 [weighted 10.7%] Hispanic, 537 [weighted 52.7%] non-Hispanic Black, 324 [weighted 32.2%] non-Hispanic White, and 42 [weighted 4.4%] of non-Hispanic other race or ethnicity were included in this study. A total of 274 patients (weighted 26.6%) met criteria for 3-month PTSD or MDE. An ensemble machine learning model restricted to 30 predictors estimated in a training sample (patients from the Northeast or Midwest) had good prediction accuracy (mean [SE] area under the curve [AUC], 0.815 [0.031]) and calibration (mean [SE] integrated calibration index, 0.040 [0.002]; mean [SE] expected calibration error, 0.039 [0.002]) in an independent test sample (patients from the South). Patients in the top 30% of predicted risk accounted for 65% of all 3-month PTSD or MDE, with a mean (SE) positive predictive value of 58.2% (6.4%) among these patients at high risk. The model had good consistency across regions of the country in terms of both AUC (mean [SE], 0.789 [0.025] using the Northeast as the test sample and 0.809 [0.023] using the Midwest as the test sample) and calibration (mean [SE] integrated calibration index, 0.048 [0.003] using the Northeast as the test sample and 0.024 [0.001] using the Midwest as the test sample; mean [SE] expected calibration error, 0.034 [0.003] using the Northeast as the test sample and 0.025 [0.001] using the Midwest as the test sample). The most important predictors in terms of Shapley Additive Explanations values were symptoms of anxiety sensitivity and depressive disposition, psychological distress in the 30 days before motor vehicle collision, and peritraumatic psychosomatic symptoms.

CONCLUSIONS AND RELEVANCE The results of this study suggest that a short set of questions feasible to administer in an ED can predict 3-month PTSD or MDE with good AUC, calibration, and geographic consistency. Patients at high risk can be identified in the ED for targeting if cost-effective preventive interventions are developed.

JAMA Psychiatry. 2021;78(11):1228-1237. doi:10.1001/jamapsychiatry.2021.2427
Published online September 1, 2021.

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: The AURORA Consortium authors appear at the end of the article.

Corresponding Author: Ronald C. Kessler, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA 02115 (kessler@hcp.med.harvard.edu).

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

Adverse posttraumatic neuropsychiatric sequelae (APNS) of traumatic experiences have a substantial societal burden.[1,2] Although posttraumatic stress disorder (PTSD) is the most frequently studied APNS, major depressive episode (MDE) is also common.[3,4] Many people who develop these APNS are evaluated in emergency departments (EDs) shortly after their traumas,[5-7] making preventive interventions possible.[8] Although theory and some preliminary empirical studies suggest that certain preventive interventions might be effective for at least some of these patients,[6] this area of research is underdeveloped. Although even before developing and evaluating preventive interventions, it would be useful to know how well we can pinpoint patients at high risk among the 40 million Americans who present annually to EDs after a trauma[9] given that it would likely be cost-effective to provide preventive interventions only to patients at high risk.

Several previous studies[10-16] attempted to develop machine learning (ML) models to predict PTSD among patients presenting to EDs after traumas. These models had good accuracy in terms of area under the receiver operating characteristic curve (AUC) predicting PTSD at 3 months (AUC, 0.79-0.85)[13,14] and 12 to 15 months (AUC, 0.71)[10,16] after trauma exposure and persistent PTSD (AUC, 0.75-0.89).[10-12,15] However, all of these studies[10-16] focused on the approximately 5% of patients with trauma who were hospitalized.[17] The APNS prevalence is equally high among the 95% of patients with trauma who are discharged.[18]

We present the results of an analysis based on the Advancing Understanding of Recovery After Trauma (AURORA) study, a longitudinal study of the onset and course of APNS among patients presenting to an ED after a traumatic experience. We included patients discharged from the ED and those hospitalized up to 72 hours.[18] We focused on motor vehicle collisions (MVCs), the most common trauma in industrialized countries[19] and in AURORA. We developed a model to predict PTSD or MDE 3 months after the ED visit compared with the exclusive focus on PTSD in prior studies.[10-16] Although previous studies[10-16] were limited to data from patients in 1 or 2 EDs, we used data from patients in 28 EDs. We trained the model using data from patients in EDs in the Midwest and Northeast and tested the model using data from patients in EDs in the South. The predictors considered were a mix of observations (eg, patient sex and race/ethnicity), standard clinical evaluations (eg, injury site and severity and vital signs), and patient reports. Although previous studies[10-16] used up to 105 predictors in their models, we aimed to develop a parsimonious model with a small number of predictors that could feasibly be administered in EDs.

Key Points

Question Is it possible to predict which patients will have posttraumatic stress disorder (PTSD) or major depressive episode (MDE) 3 months after presenting to an emergency department (ED) because of a motor vehicle collision?

Findings In this cohort study of 1003 patients evaluated in 28 US EDs, a machine learning model restricted to 30 variables found good validated area under the curve and calibration in predicting 3-month PTSD or MDE. The 30% of patients with highest predicted risk accounted for 65% of all 3-month PTSD or MDE.

Meaning These results suggest that patients at high risk can be identified in the ED for targeting if cost-effective preventive interventions are developed.

Methods

Sample

AURORA enrollment began on September 25, 2017. The patients considered in this report completed 3-month assessments by January 31, 2020, at 28 urban EDs across 3 US regions (Midwest, Northeast, and South). Patients had to be 18 to 75 years of age, presenting within 72 hours of the MVC, able to speak and read English, oriented to time and place, able to comprehend the enrollment protocol, and possessing a smartphone for more than 1 year. We excluded patients with a solid organ injury of grade 1 or higher, significant hemorrhage, or need for a chest tube or operation with general anesthesia. We initially excluded patients likely to be admitted but subsequently relaxed that exclusion to include patients admitted for no more than 24 hours (as of April 4, 2018) and then for no more than 72 hours (as of December 11, 2018). A predictor variable distinguishing those admitted vs discharged was included in the analysis. All participants provided written informed consent. All data were deidentified. This study was approved by institutional review boards at each participating institution. The study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline[20] for reporting analyses designed to develop and validate predictive models.

Patients self-reported their race by selecting one or more of the following categories: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islander, White, or any other race. To assess Hispanic ethnicity, patients were asked "Do you consider yourself to be of Hispanic, Latino, or Spanish origin." Using these two variables, we created a race and ethnicity variable with 4 categories, assigned in the following order: Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic other.

Each patient received an interviewer-administered ED assessment with self-report questions and biological sample collections described elsewhere.[18] Subsequent 2-week, 8-week, and 3-month web surveys were sent by text or email for self-completion or with telephone interviewer assistance. Patients were reimbursed $60 for the ED assessment and $40 for each follow-up survey. Of the 2096 patients presenting after an MVC and completing the baseline assessment, 1003 completed all 3 follow-up surveys (eFigure in the Supplement). We focus on these 1003 patients. An inverse probability weight was used to adjust for differences in baseline measures between these 1003 patients and those in the baseline sample who missed at least 1 follow-up.[21]

Measures

We included 394 potential predictors that spanned 11 broad APNS risk factor domains that included MVC characteristics,

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

peritraumatic signs and symptoms, chronic stressors, prior lifetime traumas, past 30-day psychological distress, physical health, past 30-day role impairment, lifetime mental disorders, sociodemographic characteristics, social support, and personality. A detailed list of constructs, measures, and scoring rules is presented in eTable 1 in the Supplement. Categorical variables were dummy coded. Quantitative variables were standardized to a mean of 0 and variance of 1 for use in linear algorithms and transformed into deciles for use in tree-based algorithms.

Outcome

The outcome for the prediction model was self-reported PTSD or MDE during a 30-day recall period assessed in the 3-month survey. Posttraumatic stress disorder was assessed with the 20-item PTSD Checklist for *DSM-5*[22] Of the several diagnostic classification rules proposed for the PTSD Checklist,[23,24] we selected a conservative threshold of 38 or higher. Major depressive episode was assessed with the Patient Reported Outcomes Measurement Information System Depression Short-Form 8b.[25] Patients were classified as meeting 3-month criteria for MDE if their scores were 30 or higher, which is 1.65 SDs above the established general population mean based on the conservative assumption that 5% of the general population meets the criteria for MDE. The outcome was defined as positive if the patient met 3-month criteria for PTSD and/or MDE. We also assessed 3-month role impairment using a modified version of the Sheehan Disability Scale[26] and the World Health Organization Disability Assessment Schedule[27] question about days totally out of role because of health problems in the past 30 days (eTable 1 in the Supplement).

Statistical Analysis

Data analysis was performed from November 1, 2020, to May 31, 2021. Patients had to complete all 3 follow-up assessments (2-week, 8-week, and 3-month assessments) because some predictors, although referring to experiences or patient characteristics before the MVC, were assessed in the 2-week or 8-week surveys to reduce patient burden in the ED. We treated the 2- and 8-week measures as if assessed at baseline. We used mean imputation for the small amount of item-missing data. To account for potential selection bias from nonresponse in follow-up surveys, we used inverse probability of response weights to adjust for the modest differences found between baseline characteristics of patients in the analysis sample and patients who did not complete at least 1 follow-up assessment.[21] All analyses were performed in this weighted data set. Weighted means of baseline variables in the analysis sample were all within 0.1 SD of the means in the total baseline sample (eTable 2 in the Supplement).

Substantive analysis began by comparing prevalence, comorbidity, and role impairments of PTSD and MDE at 3 months using 2-sided χ^2 and F tests. We then developed an ML model to predict 3-month PTSD or MDE from the baseline variables. We used a stacked generalization method in which results were pooled across multiple algorithms by generating an algorithm weight via 10-fold cross-validation in a training sample for each algorithm in the set we used (ensemble). The composite predicted outcome score is guaranteed in expectation to perform at least as well as the best component algorithm according to a prespecified criterion, which we defined as AUC.[28] The Super Learner ensemble ML method was used to implement this analysis.[29] Consistent with recommendations,[30,31] we used a diverse set of algorithms in the Super Learner ensemble to capture nonlinearities and interactions and reduce risk of misspecification.[32] These algorithms included several different linear algorithms (logistic regression, regularized regression, spline and polynomial spline regressions, and support vector machines) and regression tree-based algorithms (boosting and bagging ensemble trees and bayesian additive regression trees) (eTable 3 in the Supplement). Broadly similar stacking approaches have been used in prior ED research on PTSD[15] as well as in other computational psychiatric research studies.[33,34] Given the small sample size, hyperparameter tuning was achieved by including individual algorithms multiple times in the ensemble with different hyperparameter values and allowing Super Learner to weight relative importance across this range rather than using an external grid search or random search procedure.

Feature selection was performed independently in each 10-fold cross-validation training sample. We explored 2 different feature reduction methods, least absolute shrinkage and selection operator (LASSO) penalized regression[35] and random forest,[36] to increase feasibility of implementation in clinical practice and to reduce overfitting. The training sample was defined as the 784 patients in the Northeast or Midwest and the test sample as the 219 patients in the South. Model fit across specifications was evaluated in the test sample based on AUC. Once a best-model specification was determined, we used a locally estimated scatterplot smoothed calibration curve[37] to quantify calibration of predicted outcome probabilities using the integrated calibration index (ICI) and expected calibration error (ECE).[38,39] We additionally examined how the best-model specification would perform in terms of AUC and calibration in alternative test samples (ie, if the test samples were instead the Northeast or Midwest). We then divided the test sample into 20 ventiles of predicted risk defined in the training sample and calculated conditional and cumulative sensitivity (the proportion of patients with the outcome) and positive predictive value (PPV; prevalence of the outcome) in the test sample within and across these predicted risk ventiles. Model fairness, defined as whether model performance was comparable across important segments of the population,[40] was examined by estimating variation in the association of predicted risk with the observed outcome across subgroups defined by several key patient sociodemographic characteristics (age, sex, race/ethnicity, and income) using a robust Poisson regression model.[41] We examined predictor importance with the model-agnostic Kernel SHAP (Shapley Additive Explanations) method, which estimates the marginal contribution to overall model accuracy of each variable in a predictor set.[42] A 2-sided $P < .05$ was considered to be statistically significant.

Data management and calculations of prevalence and AUC were performed in SAS statistical software, version 9.4 (SAS Institute Inc).[43] The Super Learner models were estimated in R, version 3.6.3 (R Foundation for Statistical Computing).[44]

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

SHAP values were estimated in Python, version 3.8.5 (Python Software Foundation).[45] The R packages used for each algorithm are listed in eTable 3 in the Supplement.

Results

Prevalence of 3-Month PTSD or MDE

A total of 1003 patients (median [interquartile range] age, 34.5 [24-43] years; 715 [weighted 67.9%] female; 100 [weighted 10.7%] Hispanic, 537 [weighted 52.7%] non-Hispanic Black, 324 [weighted 32.2%] non-Hispanic White, and 42 [weighted 4.4%] of non-Hispanic other race or ethnicity were included in this study. The 3-month prevalence (SE) was 25.1% (1.4) for PTSD, 11.5% (1.0) for MDE, and 26.6% (1.4) for either (eTable 4 in the Supplement). These prevalence (SE) estimates were not markedly different from those reported retrospectively in the ED for the 30 days before MVC: 20.7% (1.3) for PTSD, 6.2% (0.8) for MDE, and 22.3% (1.3) for either. However, as noted below, our best model substantially outperformed a model using only pre-MVC PTSD and MDE to predict the 3-month outcome.

Even though 3-month MDE alone was much less common than PTSD alone (1.6% vs 15.1%; $\chi^2_1 = 11.1$; $P < .001$), the mean (SE) number of days out of role was significantly higher among patients with comorbid PTSD and MDE than among patients with PTSD alone (6.0 [0.8] vs 3.8 [0.7], $F_1 = 4.1$, $P = .04$). In addition, the mean (SE) number of days out of role was substantially higher, although not significantly so, among the small number of patients with MDE alone than those with PTSD alone (7.6 [2.9] vs 3.8 [0.7]; $F_1 = 1.6$; $P = .21$). Broadly similar results were found for patient reports of severe role impairment (eTable 5 in the Supplement). On the basis of these results, we defined our outcome as 3-month PTSD and/or MDE rather than focusing only on PTSD. The prevalence (SE) of this outcome was comparable across the 3 regions where AURORA was performed: Northeast (n [number of patients in the region] = 352; 26.5% [percentage of those patients] [2.4]), Midwest (n = 432; 26.8% [2.2]), and South (n = 219; 26.6% [3.1]).

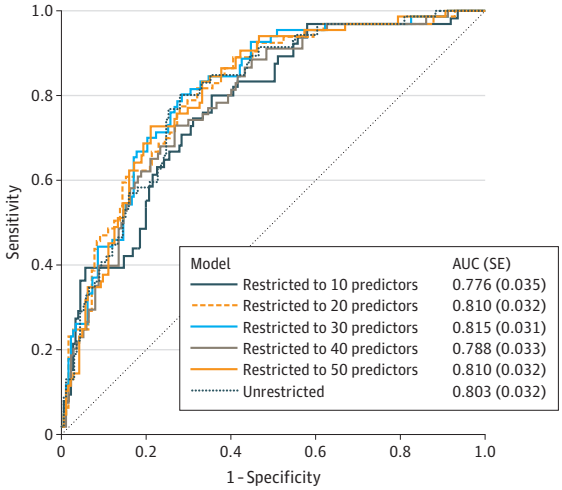
Model Performance

The mean (SE) AUC of the initial Super Learner model in the test sample was 0.803 (0.032) when only LASSO was used for feature selection and 0.782 (0.034) when both LASSO and ranger were used for feature selection. The AUC in the test sample was 0.663 (0.037), in comparison, when pre-MVC PTSD and MDE were the only predictors in a logistic regression model that allowed for interactions between these 2 predictors. On the basis of these results, we focused further analysis on restricted models that used only LASSO for feature selection and examined models restricted to 10 to 50 predictors. The AUC was higher in models restricted to 20, 30, or 50 predictors (mean [SE] AUC, 0.810 [0.032] for models with 20 predictors, 0.815 [0.031] for models with 30 predictors, and 0.810 [0.032] for models with 50 predictors) than the model with unrestricted predictors (mean [SE] AUC, 0.803 [0.032]) (**Figure 1**).

Given that the 30-predictor model had a marginally higher AUC than the others, we focused on it for further evaluation as our best model. This model had good calibration in the test sample (mean [SE] ICI, 0.040 [0.002]; mean [SE] ECE, 0.039 [0.002]). Five of the 32 algorithms in the model's ensemble accounted for almost all the Super Learner weight: 2 of the 5 extreme gradient boosting algorithms (0.32-0.38 weights), 1 of the 3 random forest algorithms (0.18 weight), and 2 of the 11 penalized logistic regression algorithms (0.01-0.11 weights) (eTable 6 in the Supplement). The mean (SE) 30-predictor model AUC in the total test sample was 0.815 (0.031). The mean (SE) AUC was 0.709 (0.067) among patients who met criteria for PTSD and/or MDE in the 30 days before MVC and 0.791 (0.046) among patients who did not meet the pre-MVC criteria for either disorder. Fairness of the model was documented by finding that the relative risk of the outcome based on predicted probabilities from the model was comparable across test sample subgroups defined by age, sex, race/ethnicity, and income (eTable 7 in the Supplement). Geographic consistency of model performance was documented by finding comparable AUC (mean [SE] AUCs, 0.789 [0.025] using the Northeast as the test sample and 0.809 [0.023] using the Midwest as the test sample) (**Figure 2**) and calibration (mean [SE] integrated calibration index, 0.048 [0.003] using the Northeast as the test sample and 0.024 [0.001] using the Midwest as the test sample; mean [SE] ECE, 0.034 [0.003] using the Northeast as the test sample and 0.025 [0.001] using the Midwest as the test sample) (**Figure 3**) when the test sample was changed to be patients in the Northeast or Midwest.

Inspection of model sensitivity and PPV found that, despite some nonmonotonicity, patients in the top 5 predicted training sample risk ventiles, which included 29.9% of the test sample, had sensitivities between 1.7 and 2.8 times the value expected by chance, whereas remaining patients had sensitivities near (ventiles 5-10) or below (ventiles 11-20) expected

Figure 1. Area Under the Receiver Operating Characteristic Curves (AUCs) in the Test Sample (n = 219) Predicting 3-Month Posttraumatic Stress Disorder or Major Depressive Episode Based on Super Learner Models With Restricted and Unrestricted Least Absolute Shrinkage and Selection Operator Feature Selection Estimated in the Training Sample (n = 784)



© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

Figure 2. Area Under the Receiver Operating Characteristic Curves (AUCs) in Alternative Test Samples Defined by Census Region Predicting 3-Month Posttraumatic Stress Disorder or Major Depressive Episode Based on Super Learner Models With Least Absolute Shrinkage and Selection Operator Feature Selection Restricted to 30 Predictors

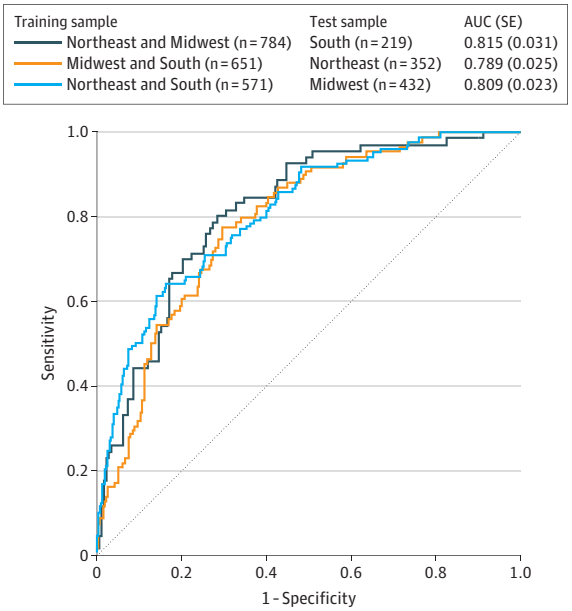
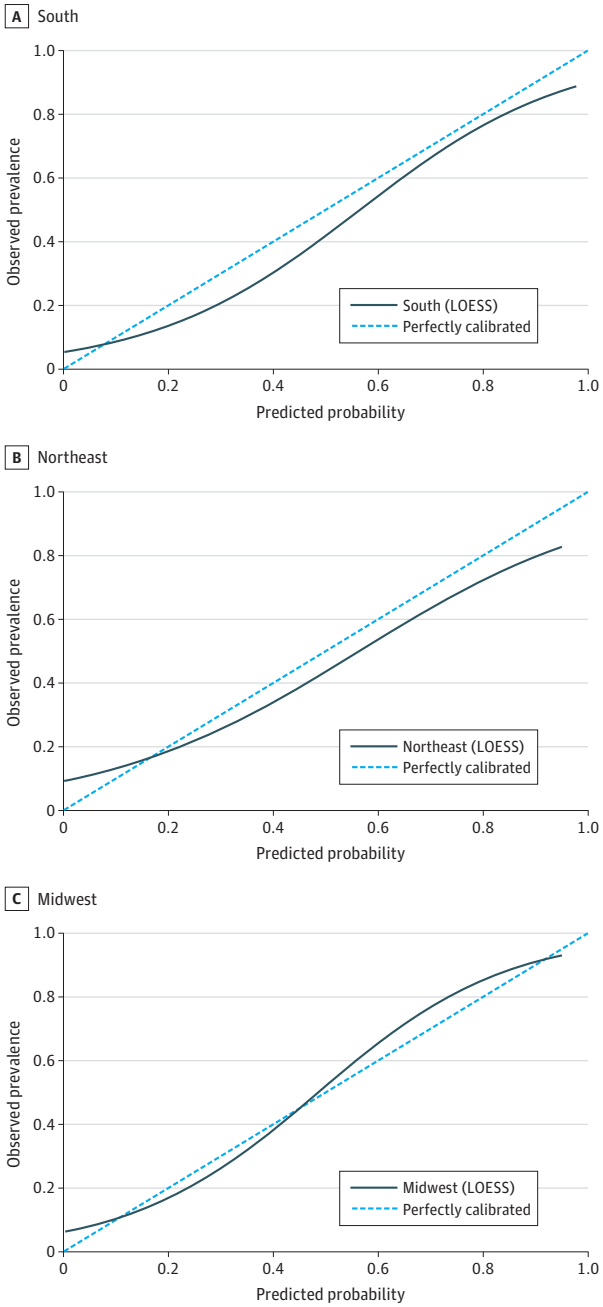


Figure 3. Locally Estimated Scatterplot Smoothing (LOESS) Calibration Curves for Predicted Probability of 3-Month Posttraumatic Stress Disorder or Major Depressive Episode in Alternative Test Samples Defined by Census Region Based on Super Learner Models With Least Absolute Shrinkage and Selection Operator Feature Selection Restricted to 30 Predictors



The smoothing span is 0.75, the integrated calibration index is 0.024 to 0.0409, and the expected calibration error is 0.025 to 0.039.

values (**Table**). Cumulative sensitivity across the top 5 ventiles was 65.4%, and the cumulative PPV in that range was 58.2%.

Predictor Importance

A total of 264 of the 394 variables (67%) in the predictor set had zero-order associations with the outcome in the total sample, including 94% to 100% of those assessing 30 days before MVC psychological distress and impairment and recent stressors; 70% to 85% of those assessing peritraumatic symptoms, social support, and personality; 50% to 60% of those assessing lifetime traumas and mental disorders and physical health; and 25% to 30% of those assessing sociodemographic and MVC characteristics (eTable 8 in the Supplement). Admission status (ie, admitted to the hospital vs discharged) was not a significant zero-order predictor (odds ratio, 1.0; 95% CI, 0.9-1.1). To examine predictor importance, we reran the best model specification (ie, 30 predictors selected by LASSO separately for linear and tree-based algorithms) in the total sample. A total of 53 predictors were selected (30 each for linear and tree-based models, with an overlap of 7 predictors), which came from 40 variables (ie, 13 were alternative transformations of the same variable) (eTable 1 in the Supplement). The 20 most important predictors accounted for 75.5% of the total mean absolute SHAP value across all predictors in the model (**Figure 4**). These predictors included 7 indicators of personality (6 of anxiety sensitivity and 1 of dispositional depression), 7 of peritraumatic psychosomatic symptoms, 4 of past 30-day psychological symptoms (2 depression, 1 PTSD, and 1 impairment attributable to emotional problems), and 2 of prior lifetime trauma exposure. The personality measures were among those assessed retrospectively in the 2-week follow-up survey. Replication of the Super Learner with LASSO feature selection of 30 predictors from a reduced predictor set that excluded retrospectively reported variables (ie, lifetime traumatic experiences) had a lower AUC in analyses sequentially treating

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

Table. Prediction of Posttraumatic Stress Disorder or Major Depressive Episode 3 Months After a Motor Vehicle Collision in the Test Sample of 219 Patients From the South by Ventiles of the Predicted Risk Distribution in the Training Sample of 784 Patients From the Northeast and Midwest Using the Super Learner Model With LASSO Feature Selection Restricted to 30 Predictors

Ventile	Total No. of patients	Within ventile			Cumulative		
		Patients, %	Sensitivity, mean (SE)	PPV, mean (SE)	Patients, %	Sensitivity, mean (SE)	PPV, mean (SE)
1	11	4.7	13.1 (12.5)	74.2 (13.1)	4.7	13.1 (4.4)	74.2 (13.1)
2	9	4.3	11.4 (11.1)	70.4 (15.4)	9.0	24.5 (6.0)	72.4 (10.0)
3	14	6.5	12.5 (12.0)	51.0 (13.7)	15.5	36.9 (6.6)	63.4 (8.5)
4	20	10.0	17.3 (15.6)	45.8 (11.4)	25.6	54.2 (6.6)	56.5 (7.0)
5	10	4.4	11.2 (10.9)	68.2 (15.4)	29.9	65.4 (6.2)	58.2 (6.4)
6	18	8.0	7.6 (7.7)	25.4 (10.2)	37.9	73.0 (5.8)	51.3 (5.7)
7	9	3.5	5.6 (5.8)	42.8 (16.6)	41.4	78.6 (5.4)	50.6 (5.4)
8	12	5.4	4.7 (4.9)	23.1 (12.0)	46.8	83.3 (5.0)	47.4 (5.1)
9	17	8.4	5.4 (5.6)	17.0 (9.5)	55.2	88.7 (4.1)	42.8 (4.7)
10	12	5.7	4.0 (4.2)	18.6 (11.8)	60.9	92.7 (3.2)	40.5 (4.4)
11	7	3.3	2.8 (3.0)	22.8 (14.9)	64.2	95.5 (2.6)	39.6 (4.3)
12	19	9.4	1.4 (1.5)	4.0 (4.0)	73.7	96.9 (2.2)	35.0 (3.9)
13	13	5.7	0	0	79.4	96.9 (2.2)	32.5 (3.7)
14	4	1.6	0	0	81.0	96.9 (2.2)	31.9 (3.6)
15	7	3.0	0	0	84.0	96.9 (2.2)	30.7 (3.5)
16	9	4.1	1.8 (1.9)	11.5 (11.0)	88.1	98.7 (1.3)	29.8 (3.4)
17	10	4.2	0	0	92.4	98.7 (1.3)	28.4 (3.3)
18	4	1.6	1.3 (1.4)	22.4 (20.7)	93.9	100.0	28.3 (3.2)
19	3	1.3	0	0	95.2	100.0	28.0 (3.2)
20	11	4.8	0	0	100.0	100.0	26.6 (3.1)

Abbreviations: LASSO, least absolute shrinkage and selection operator; PPV, positive predictive value.

patients in 2 regions as the training sample and those in the third region as the test sample (AUC [SE], 0.815 [0.031] using the South as the test sample, 0.789 [0.025] using the Northeast as the test sample, and 0.809 [0.023] using the Midwest as the test sample) than when the retrospectively reported variables were included (AUC [SE], 0.755 [0.035] using the South as the test sample, 0.748 [0.031] using the Northeast as the test sample, and 0.754 [0.027] using the Midwest as the test sample).

Discussion

In this study, our model's AUC was comparable to models developed in previous ED studies to predict persistent PTSD,[10-12,15] 3-month PTSD,[13,14] or 12- to 15-month PTSD.[10,16] However, these other studies[10-16] used up to 105 predictors vs 40 in our model, and many of the most important predictors in prior studies[15,16] were laboratory tests that are routinely performed only for patients with trauma admitted to the hospital, which do not apply to the approximately 95% of ED patients discharged to home. The external validity of earlier models was also limited by their inclusion of only 1 or 2 EDs. In addition, whereas our model was well calibrated, only 1 previous study[15] examined calibration and found it to be relatively poor.

Caution is needed in interpreting our findings regarding predictor importance because this depends on associations of predictors with each other. It is nonetheless interesting that items assessing dispositional anxiety sensitivity emerged as the most important predictors. Such measures were not included in previous studies.[10-16] The other 2 most important predictor domains in our model were peritraumatic psychosomatic symptoms in the ED and psychological distress. Only 2 prior studies assessed psychological distress in the weeks[13] or months[14] before trauma exposure. Both found that these were important predictors. Although no prior study assessed peritraumatic psychosomatic symptoms, some assessed peritraumatic distress[10,12,14,15] and dissociation[14,15] and found both to be important predictors. Consistent with these prior results, we found that peritraumatic distress and dissociation were significant univariate predictors of our outcome, although they were not selected in the final model.

It is also important to recognize that the value of our model depends on unknowns about the costs and effects of preventive interventions. As noted above, this is an underdeveloped area of research.[6] Determining whether the PPV of our model at a decision threshold is sufficiently high to justify implementing a targeted intervention would, at a minimum, require an evaluation of the precision recall curve and, importantly, the net benefit curve[46] based on a formal cost-effectiveness analysis. In addition, if heterogeneity of treatment effects is found, the development of an individualized precision treatment rule would be required to evaluate the effects of our prediction model.[47]

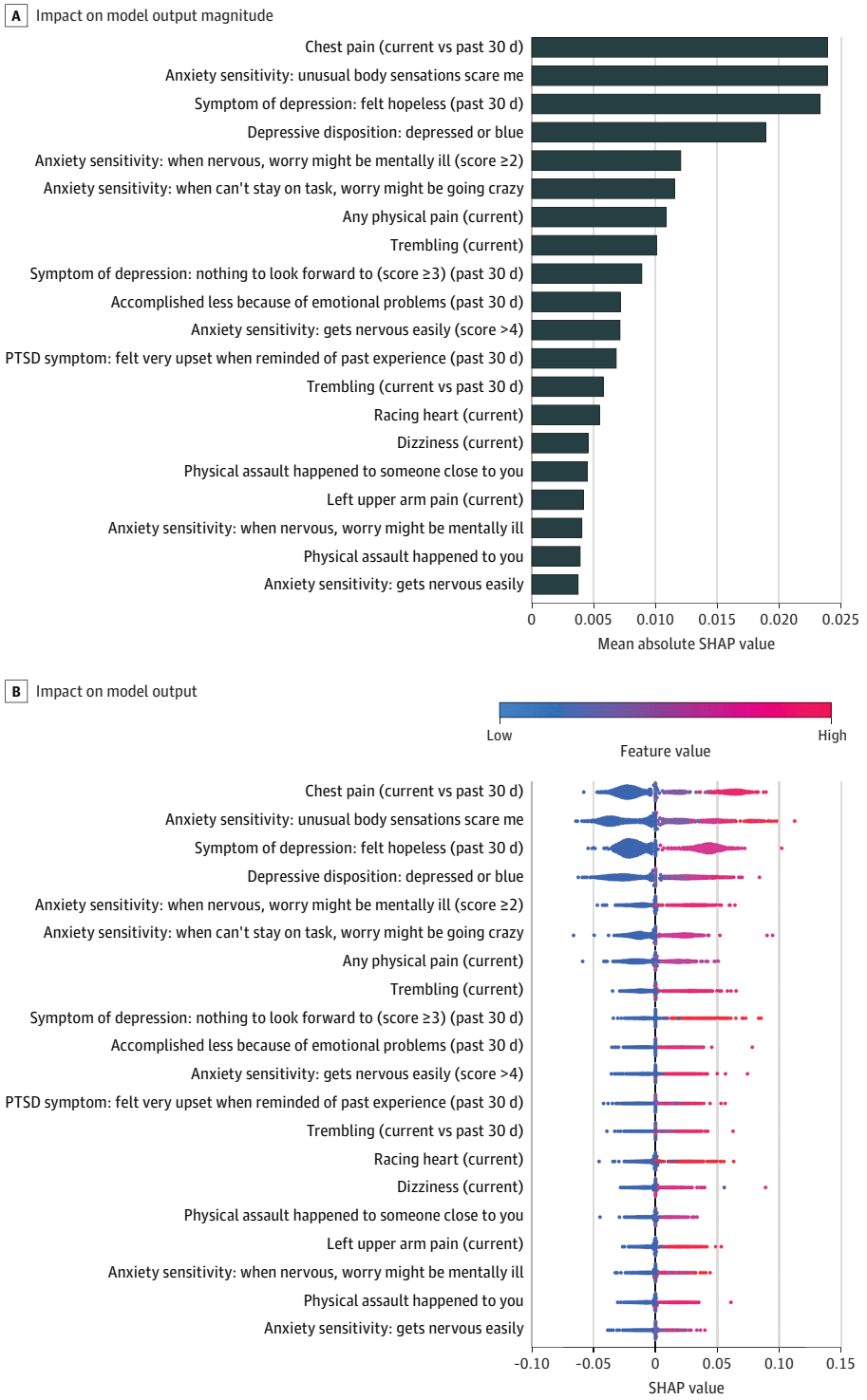
Limitations

Our study has several noteworthy limitations. First, the sample included only English-speaking patients from urban EDs after

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

Figure 4. Predictor Importance Determined by Shapley Additive Explanations (SHAP) Values for the Super Learner Model With Least Absolute Shrinkage and Selection Operator Feature Selection Restricted to 30 Predictors



The bar chart shows the mean absolute SHAP values for the top 20 predictors. Larger SHAP values indicate greater importance for predicting 3-month posttraumatic stress disorder (PTSD) or major depressive episode. The SHAP heat map displays the SHAP values for each of the top 20 predictors sorted by their global importance (with more important predictors at the top). Each dot represents an individual's SHAP value for that predictor. The value of each individual's predictor ranges from low (blue) to high (red), and its corresponding SHAP value (the mean absolute value change in predicted probability of the outcome associated with the variable) is also shown.

an MVC who were followed up for 3 months. Different samples and follow-up periods might yield different results. Second, the response rate was low, raising the possibility of sample selection bias. Third, patients with pre-MVC PTSD and MDE were not excluded, although our AUC was substantially higher than in a model in which 30-day pre-MVC PTSD and MDE were the only predictors, and only 3 of our top 20 predictors were symptoms of 30-day pre-MVC PTSD or MDE. Fourth, we did not consider

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

the small number of patients who were hospitalized for more than 72 hours. We also did not obtain information about outpatient treatment after ED discharge. These omissions could have reduced the external validity by excluding otherwise important baseline variables with effects on 3-month outcomes mediated by treatment. Fifth, outcome measures were based on validated self-report scales rather than clinical interviews.[22,25] Sixth, some important predictors were assessed in the 2-week surveys, and overall model prediction accuracy was lower when these variables were omitted from the model. Replication in a sample that assesses these variables at baseline will be needed to determine their true importance.

Conclusion

This study found that a parsimonious model that predicts 3-month PTSD or MDE after MVC can be developed using a battery of questions that could be delivered in approximately 10 minutes. The model had good AUC and calibration and captured close to two-thirds of all patients who developed 3-month PTSD or MDE in the top 30% of the predicted risk distribution. These results suggest that if cost-effective preventive interventions are developed, identification of patients in the ED who are at high risk for treatment targeting may be possible.

ARTICLE INFORMATION

Accepted for Publication: June 30, 2021.

Published Online: September 1, 2021. doi:10.1001/jamapsychiatry.2021.2427

Author Affiliations: Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Ziobrowski, Petukhova, Sampson, Puac-Polanco, Lee, Kessler); Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts (Kennedy); Halıcıoğlu Data Science Institute, University of California, San Diego (Ustun); Department of Emergency Medicine, Washington University School of Medicine, St Louis, Missouri (House); Department of Emergency Medicine & Department of Health Services, Policy, and Practice, The Alpert Medical School of Brown University, Rhode Island Hospital and The Miriam Hospital, Providence, Rhode Island (Beaudoin); Institute for Trauma Recovery, Department of Anesthesiology, University of North Carolina at Chapel Hill (An, McLean); Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill (Zeng); Department of Psychology and Neuroscience & Department of Sociology, University of North Carolina at Chapel Hill (Bollen); Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York (Puac-Polanco); Department of Epidemiology, Harvard T. H. Chan School of Public Health, Harvard University, Boston, Massachusetts (Koenen); Department of Psychiatry, Harvard Medical School, Boston, Massachusetts (Ressler); Division of Depression and Anxiety, McLean Hospital, Belmont, Massachusetts (Ressler); Department of Emergency Medicine, University of North Carolina at Chapel Hill (McLean).

The AURORA Consortium authors: Jennifer S. Stevens, PhD; Thomas C. Neylan, MD; Gari D. Clifford, DPhil; Tanja Jovanovic, PhD; Sarah D. Linnstaedt, PhD; Laura T. Germine, PhD; Scott L. Rauch, MD; John P. Haran, MD, PhD; Alan B. Storrow, MD; Christopher Lewandowski, MD; Paul I. Musey Jr, MD; Phyllis L. Hendry, MD; Sophia Sheikh, MD; Christopher W. Jones, MD; Brittany E. Punches, PhD, RN; Michael S. Lyons, MD, MPH; Vishnu P. Murty, PhD; Meghan E. McGrath, MD; Jose L. Pascual, MD, PhD; Mark J. Seamon, MD; Elizabeth M. Datner, MD; Anna M. Chang, MD; Claire Pearson, MD; David A. Peak, MD; Guruprasad Jambaulikar, MBBS, MPH; Roland C. Merchant, MD, ScD, MPH; Robert M. Domeier, MD; Niels K. Rathlev, MD; Brian J. O'Neil, MD; Paulina Sergot, MD; Leon D. Sanchez, MD, MPH; Steven E. Bruce, PhD; Robert H. Pietrzak, PhD, MPH; Jutta Joormann, PhD; Deanna M. Barch, PhD; Diego A. Pizzagalli, PhD; John F. Sheridan, PhD; Steven E. Harte, PhD; James M. Elliott, PhD; Sanne J. H. van Rooij, PhD.

Affiliations of The AURORA Consortium authors: Institute for Trauma Recovery, Department of Anesthesiology, University of North Carolina at Chapel Hill (Linnstaedt); Department of Psychiatry, Harvard Medical School, Boston, Massachusetts (Germine, Rauch, Pizzagalli); Division of Depression and Anxiety, McLean Hospital, Belmont, Massachusetts (Pizzagalli); Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia (Stevens, van Rooij); Departments of Psychiatry and Neurology, University of California, San Francisco (Neylan); Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, Georgia (Clifford); Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta (Clifford); Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, Michigan (Jovanovic); Institute for Technology in Psychiatry, McLean Hospital, Belmont, Massachusetts (Germine, Rauch); The Many Brains Project, Belmont, Massachusetts (Germine); Department of Psychiatry, McLean Hospital, Belmont, Massachusetts (Rauch); Department of Emergency Medicine, University of Massachusetts Medical School, Worcester (Haran); Department of Emergency Medicine, Vanderbilt University Medical Center, Nashville, Tennessee (Storrow); Department of Emergency Medicine, Henry Ford Health System, Detroit, Michigan (Lewandowski); Department of Emergency Medicine, Indiana University School of Medicine, Indianapolis (Musey Jr); Department of Emergency Medicine, University of Florida College of Medicine, Jacksonville (Hendry, Sheikh); Department of Emergency Medicine, Cooper Medical School of Rowan University, Camden, New Jersey (Jones); Department of Emergency Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio (Punches, Lyons); College of Nursing, University of Cincinnati, Cincinnati, Ohio (Punches); Center for Addiction Research, University of Cincinnati College of Medicine, Cincinnati, Ohio (Punches, Lyons); Department of Psychology, Temple University, Philadelphia, Pennsylvania (Murty); Department of Emergency Medicine, Boston Medical Center, Boston, Massachusetts (McGrath); Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia (Pascual, Seamon); Department of Neurosurgery, University of Pennsylvania Perelman School of Medicine, Philadelphia (Pascual); Department of Emergency Medicine, Einstein Healthcare Network, Philadelphia, Pennsylvania (Datner); Department of Emergency Medicine, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania (Datner); Department of Emergency Medicine, Jefferson University Hospitals, Philadelphia, Pennsylvania (Chang); Department of Emergency Medicine, Wayne State University, Detroit, Michigan (Pearson, O'Neil); Department of Emergency Medicine, Massachusetts General Hospital, Boston (Peak); Department of Emergency Medicine, Brigham and Women's Hospital, Boston, Massachusetts (Jambaulikar, Merchant); Department of Emergency Medicine, Saint Joseph Mercy Hospital, Ypsilanti, Michigan (Domeier); Department of Emergency Medicine, University of Massachusetts Medical School–Baystate, Springfield (Rathlev); McGovern Medical School, University of Texas Health Science Center, Houston (Sergot); Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts (Sanchez); Department of Emergency Medicine, Harvard Medical School, Boston, Massachusetts (Sanchez); Department of Psychological Sciences, University of Missouri, St Louis (Bruce); National Center for PTSD, Clinical Neurosciences Division, Veterans Affairs Connecticut Healthcare System, West Haven (Pietrzak); Department of Psychiatry, Yale School of Medicine, West Haven, Connecticut (Pietrzak); Department of Psychology, Yale University, West Haven, Connecticut (Joormann); Department of Psychological & Brain Sciences, Washington University, St Louis, Missouri (Barch); Center for Depression, Anxiety, and Stress Research, McLean Hospital, Belmont, Massachusetts (Pizzagalli); Department of Biosciences and Neuroscience, Wexner Medical Center, The Ohio State University, Columbus (Sheridan); Institute for Behavioral Medicine Research, Wexner Medical Center, The Ohio State University, Columbus (Sheridan); Department of Anesthesiology, University of Michigan Medical School, Ann Arbor (Harte); Department of Internal Medicine-Rheumatology, University of Michigan Medical School, Ann Arbor (Harte); Kolling Institute of Medical Research, University of Sydney, St Leonards, New South Wales, Australia (Elliott); Faculty of Medicine and Health, University of Sydney, Northern Sydney Local Health District, New South Wales, Australia (Elliott); Department of Physical Therapy & Human Movement Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois (Elliott).

Author Contributions: Dr Kessler had full access to all the data in the study and takes responsibility for

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

the integrity of the data and the accuracy of the data analysis.
Concept and design: Ziobrowski, Bollen, Koenen, Ressler, McLean, Kessler, Stevens, Neylan, Clifford, Jovanovic, Germine, Rauch, Murty, McGrath, Peak, Rathlev, Joormann, Barch, Pizzagalli, Sheridan, Harte, Elliott, van Rooij.
Acquisition, analysis, or interpretation of data: Ziobrowski, Kennedy, Ustun, House, Beaudoin, An, Zeng, Petukhova, Sampson, Puac-Polanco, Lee, Koenen, McLean, Kessler, Neylan, Clifford, Jovanovic, Linnstaedt, Rauch, Haran, Storrow, Lewandowski, Musey Jr, Hendry, Sheikh, Jones, Punches, Lyons, McGrath, Pascual, Seamon, Datner, Chang, Pearson, Peak, Jambaulikar, Merchant, Domeier, O'Neil, Sergot, Sanchez, Bruce, Pietrzak, Joormann, Barch, Harte.
Drafting of the manuscript: Ziobrowski, McLean, Kessler, Haran, Rathlev, Joormann, Elliott.
Critical revision of the manuscript for important intellectual content: Ziobrowski, Kennedy, Ustun, House, Beaudoin, An, Zeng, Bollen, Petukhova, Sampson, Puac-Polanco, Lee, Koenen, Ressler, McLean, Kessler, Stevens, Neylan, Clifford, Jovanovic, Linnstaedt, Germine, Rauch, Storrow, Lewandowski, Musey Jr, Hendry, Sheikh, Jones, Punches, Lyons, Murty, McGrath, Pascual, Seamon, Datner, Chang, Pearson, Peak, Jambaulikar, Merchant, Domeier, Rathlev, O'Neil, Sergot, Sanchez, Bruce, Pietrzak, Joormann, Barch, Pizzagalli, Sheridan, Harte, Elliott, van Rooij.
Statistical analysis: Ziobrowski, Kennedy, Ustun, An, Petukhova, Sampson, Lee.
Obtained funding: Bollen, Koenen, Ressler, McLean, Kessler, Neylan, Germine.
Administrative, technical, or material support: Ziobrowski, Beaudoin, Puac-Polanco, Ressler, McLean, Stevens, Neylan, Clifford, Linnstaedt, Germine, Storrow, Lewandowski, Hendry, Sheikh, Punches, Murty, Datner, Chang, Peak, Jambaulikar, Rathlev, O'Neil, Sergot, Sanchez, Bruce, Pietrzak, Barch, Pizzagalli, Harte, van Rooij.
Supervision: Sampson, Ressler, McLean, Kessler, Jovanovic, Haran, Storrow, Lewandowski, Seamon, Chang, Jambaulikar, Bruce, Harte.

Conflict of Interest Disclosures: Dr Ziobrowski reported receiving grants from National Institute of Mental Health (NIMH) during the conduct of the study. Dr An reported receiving grants from the NIMH, US Army Medical Research and Material Command, The One Mind Foundation, and The Mayday Fund and nonfinancial technical support in collecting and processing smartphone and smartwatch data from Verily Life Science and Mindstrong Health during the conduct of the study. Dr Sampson reported receiving grants from the NIMH during the conduct of the study. Dr Lee reported receiving grants from the NIMH during the conduct of the study. Dr Ressler reported receiving grants from Takeda and Brainsway and personal fees from Janssen, Verily, Alto Neuroscience, and Bioxcel outside the submitted work. Dr McLean reported receiving grants from the NIMH, Mindstrong Health, and Verily Life Sciences during the conduct of the study. Dr Kessler reported receiving grants from the NIMH, receiving consulting fees from DataStat Inc and Sage Pharmaceuticals, and owning stock in Mirah, PYM, and Roga Sciences during the conduct of the study. Dr Clifford reported receiving grants from University of North Carolina as a subcontract on the parent AURORA grant funding during the conduct of the study and in the past 3 years has received research funding from the National Science Foundation, National Institutes of Health (NIH), and LifeBell AI and unrestricted donations from AliveCor, Amazon Research, Center for Discovery, the Gordon and Betty Moore Foundation, MathWorks, Microsoft Research, Gates Foundation, Google, One Mind Foundation, and Samsung Research. Dr Clifford also has financial interest in AliveCor and receives unrestricted funding from the company and is the chief technical officer of MindChild Medical and the chief security officer of LifeBell AI and has ownership in both companies. Dr Jovanovic reported receiving grants from NIH during the conduct of the study and outside the submitted work. Dr Germine reported serving on the Scientific Advisory Board for Sage Bionetworks for which she receives a small honorarium. Dr Rauch reported receiving grants from NIH during the conduct of the study and grants from NIH, personal fees from Society of Biological Psychiatry, royalties from Oxford University Press and APP, a per diem for serving on the oversight committee of the Veterans Affairs, funds for board service from Community Psychiatry, including equity outside the submitted work, and having leadership roles on boards or councils for Society of Biological Psychiatry, Anxiety and Depression Association of America, and National Network of Depression Centers outside the submitted work. Dr Storrow reported receiving grants from NIH during the conduct of the study. Dr Sheikh reported receiving grants from Florida Medical Malpractice Joint Underwriter's Association, Substance Abuse and Mental Health Services Administration, Florida Blue Foundation, and NIH/National Institute on Aging–funded Jacksonville Aging Studies Center outside the submitted work. Dr Jones reported receiving grants from NIMH during the conduct of the study and grants from Vapotherm Inc, Janssen, AstraZeneca, and Hologic Inc outside the submitted work. Dr Lyons reported receiving grants from NIH during the conduct of the study. Dr Pascual reported receiving grants from Grifols SA and personal fees for expert testimony outside the submitted work. Dr Chang reported receiving grants from NIH during the conduct of the study and personal fees from Roche and grants from Abbott, Ortho Clinical Diagnostics, and Siemens outside the submitted work. Dr Pearson reported receiving grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases during the conduct of the study. Dr Bruce reported receiving grants from NIMH during the conduct of the study. Dr Joormann reported receiving personal fees from Janssen Pharmaceuticals outside the submitted work. Dr Barch reported receiving grants from National Institute of Drug Abuse and NIMH during the conduct of the study. Dr Pizzagalli reported receiving personal fees from BlackThorn Therapeutics, Boehringer Ingelheim, Compass Pathways, Concert Pharmaceuticals, Engrail Therapeutics, Neurocrine Biosciences, Otsuka Pharmaceuticals, Takeda Pharmaceuticals, and Alkermes, receiving grants from Millennium Pharmaceuticals, NIMH, Brain and Behavior Research Foundation, and Dana Foundation, and having stock options in BlackThorn Therapeutics outside the submitted work. Dr Harte reported receiving grants from Aptinyx, Arbor Medical Innovations, and NIH and personal fees from Eli Lilly outside the submitted work. Dr Elliott reported receiving personal fees from Orofacial Therapeutics Honorarium outside the submitted work. No other disclosures were reported.

Funding/Support: Advancing Understanding of Recovery After Trauma (AURORA) is supported by grant U01MH110925 from the NIMH, the US Army Medical Research and Material Command, the One Mind Foundation, and The Mayday Fund. Verily Life Sciences and Mindstrong Health provided some of the hardware and software used to perform study assessments. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Atwoli L, Stein DJ, Koenen KC, McLaughlin KA. Epidemiology of posttraumatic stress disorder: prevalence, correlates and consequences. *Curr Opin Psychiatry*. 2015;28(4):307-311. doi:10.1097/YCO.0000000000000167

2. Koenen KC, Ratanatharathorn A, Ng L, et al. Posttraumatic stress disorder in the World Mental Health surveys. *Psychol Med*. 2017;47(13):2260-2274. doi:10.1017/S0033291717000708

3. Au TM, Dickstein BD, Comer JS, Salters-Pedneault K, Litz BT. Co-occurring posttraumatic stress and depression symptoms after sexual assault: a latent profile analysis. *J Affect Disord*. 2013;149(1-3):209-216. doi:10.1016/j.jad.2013.01.026

4. Norman SB, Trim RS, Goldsmith AA, et al. Role of risk factors proximate to time of trauma in the course of PTSD and MDD symptoms following traumatic injury. *J Trauma Stress*. 2011;24(4):390-398. doi:10.1002/jts.20669

5. Birur B, Moore NC, Davis LL. An evidence-based review of early intervention and prevention of posttraumatic stress disorder. *Community Ment Health J*. 2017;53(2):183-201. doi:10.1007/s10597-016-0047-x

6. Qi W, Gevonden M, Shalev A. Prevention of post-traumatic stress disorder after trauma: current evidence and future directions. *Curr Psychiatry Rep*. 2016;18(2):20. doi:10.1007/s11920-015-0655-0

7. Shalev AY, Ankri Y, Israeli-Shalev Y, Peleg T, Adessky R, Freedman S. Prevention of posttraumatic stress disorder by early treatment: results from the Jerusalem Trauma Outreach and Prevention Study. *Arch Gen Psychiatry*. 2012;69(2):166-176. doi:10.1001/archgenpsychiatry.2011.127

8. Linares IM, Corchs FDAF, Chagas MHN, Zuardi AW, Martin-Santos R, Crippa JAS. Early interventions for the prevention of PTSD in adults: a systematic literature review. *Arch Clin Psychiatry*. 2017;44(1):23-29. doi:10.1590/0101-60830000000109

9. Centers for Disease Control and Prevention. Emergency department visits. 2021. Accessed February 9, 2021. https://www.cdc.gov/nchs/fastats/emergency-department.htm

10. Galatzer-Levy IR, Karstoft KI, Statnikov A, Shalev AY. Quantitative forecasting of PTSD from early trauma responses: a machine learning

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021

application. *J Psychiatr Res*. 2014;59:68-76. doi:10.1016/j.jpsychires.2014.08.017

11. Galatzer-Levy IR, Ma S, Statnikov A, Yehuda R, Shalev AY. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Transl Psychiatry*. 2017;7(3):e0. doi:10.1038/tp.2017.38

12. Karstoft KI, Galatzer-Levy IR, Statnikov A, et al; Jerusalem Trauma Outreach and Prevention Study (J-TOPS) group. Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC Psychiatry*. 2015;15:30. doi:10.1186/s12888-015-0399-8

13. Papini S, Pisner D, Shumake J, et al. Ensemble machine learning prediction of posttraumatic stress disorder screening status after emergency room hospitalization. *J Anxiety Disord*. 2018;60:35-42. doi:10.1016/j.janxdis.2018.10.004

14. Saxe GN, Ma S, Ren J, Aliferis C. Machine learning methods to predict child posttraumatic stress: a proof of concept study. *BMC Psychiatry*. 2017;17(1):223. doi:10.1186/s12888-017-1384-1

15. Schultebraucks K, Shalev AY, Michopoulos V, et al. A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nat Med*. 2020;26(7):1084-1088. doi:10.1038/s41591-020-0951-z

16. Schultebraucks K, Sijbrandij M, Galatzer-Levy I, Mouthaan J, Olff M, van Zuiden M. Forecasting individual risk for long-term posttraumatic stress disorder in emergency medical settings using biomedical data: a machine learning multicenter cohort study. *Neurobiol Stress*. 2021;14:100297. doi:10.1016/j.ynstr.2021.100297

17. Centers for Disease Control and Prevention. Ambulatory health care data. 2021. Accessed February 9, 2021. https://www.cdc.gov/nchs/ahcd/index.htm

18. McLean SA, Ressler K, Koenen KC, et al. The AURORA Study: a longitudinal, multimodal library of brain biology and function after traumatic stress exposure. *Mol Psychiatry*. 2020;25(2):283-296. doi:10.1038/s41380-019-0581-3

19. Benjet C, Bromet E, Karam EG, et al. The epidemiology of traumatic event exposure worldwide: results from the World Mental Health Survey Consortium. *Psychol Med*. 2016;46(2):327-343. doi:10.1017/S0033291715001981

20. Collins GS, Reitsma JB, Altman DG, et al; TRIPOD Group. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Circulation*. 2015;131(2):211-219. doi:10.1161/CIRCULATIONAHA.114.014508

21. Mansournia MA, Altman DG. Inverse probability weighting. *BMJ*. 2016;352:i189. doi:10.1136/bmj.i189

22. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. *J Trauma Stress*. 2015;28(6):489-498. doi:10.1002/jts.22059

23. Bovin MJ, Marx BP, Weathers FW, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. *Psychol Assess*. 2016;28(11):1379-1391. doi:10.1037/pas0000254

24. Zuromski KL, Ustun B, Hwang I, et al. Developing an optimal short-form of the PTSD Checklist for DSM-5 (PCL-5). *Depress Anxiety*. 2019;36(9):790-800. doi:10.1002/da.22942

25. Cella D, Riley W, Stone A, et al; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63(11):1179-1194. doi:10.1016/j.jclinepi.2010.04.011

26. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *Int J Psychiatry Med*. 1997;27(2):93-105. doi:10.2190/T8EM-C8YH-373N-1UWD

27. Ustün TB, Chatterji S, Kostanjsek N, et al; WHO/NIH Joint Project. Developing the World Health Organization disability assessment schedule 2.0. *Bull World Health Organ*. 2010;88(11):815-823. doi:10.2471/BLT.09.067231

28. Polley EC, Rose S, van der Laan MJ. Super learning. In: *Targeted learning: Casual Inference for Observational and Experimental Data*. Springer-Verlag New York; 2011:43-66. doi:10.1007/978-1-4419-9782-1_3

29. R Core Team. Superlearner: Super learner prediction, version 2.0-28. R Foundation for Statistical Computing. 2018. Accessed May 13, 2021. https://CRAN.R-project.org/package=SuperLearner

30. LeDell E, van der Laan MJ, Petersen M. AUC-maximizing ensembles through metalearning. *Int J Biostat*. 2016;12(1):203-218. doi:10.1515/ijb-2015-0035

31. Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. *Eur J Epidemiol*. 2018;33(5):459-464. doi:10.1007/s10654-018-0390-z

32. Kabir MF, Ludwig SA. Enhancing the performance of classification using super learning. *Data-Enabled Discov*. 2019;3(1):5. doi:10.1007/s41688-019-0030-0

33. Karrer TM, Bassett DS, Derntl B, et al. Brain-based ranking of cognitive domains to predict schizophrenia. *Hum Brain Mapp*. 2019;40(15):4487-4507. doi:10.1002/hbm.24716

34. Acion L, Kelmansky D, van der Laan M, Sahker E, Jones D, Arndt S. Use of a machine learning framework to predict substance use disorder treatment success. *PLoS One*. 2017;12(4):e0175383. doi:10.1371/journal.pone.0175383

35. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x

36. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324

37. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med*. 2014;33(3):517-535. doi:10.1002/sim.5941

38. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Stat Med*. 2019;38(21):4051-4065. doi:10.1002/sim.8281

39. Naeini MP, Cooper GF, Hauskrecht M. Obtaining well calibrated probabilities using bayesian binning. *Proc Conf AAAI Artif Intell*. 2015;2015:2901-2907.

40. Yuan M, Kumar V, Ahmad M, Teredesai A. Assessing fairness in classification parity of machine learning models in healthcare. Cornell University Library. Accessed February 7, 2021. https://arxiv.org/abs/2102.03717

41. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159(7):702-706. doi:10.1093/aje/kwh090

42. Lundberg SM, Lee S-I. Advances in neural information processing systems 30 (NIPS): a unified approach to interpreting model predictions. Cornell University Library. Accessed February 7, 2021. https://arxiv.org/abs/1705.07874

43. *SAS/STAT*. Version 9.4 for Unix. SAS Institute Inc; 2016.

44. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. 2018. Accessed May 13, 2021. https://www.R-project.org/

45. Lundberg SM. Welcome to the SHAP documentation. Accessed May 13, 2021. https://shap.readthedocs.io/en/latest/index.html

46. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J*. 2021;S1529-9430(21)00112-1. doi:10.1016/j.spinee.2021.02.024

47. Kessler RC, Furukawa TA, Kato T, et al. An individualized treatment rule to optimize probability of remission by continuation, switching, or combining antidepressant medications after failing a first-line antidepressant in a two-stage randomized trial. *Psychol Med*. 2021;1-10. doi:10.1017/S0033291721000027

© 2021 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Washington University - St Louis User on 11/22/2021